

Cílem této diplomové práce je navrhnout a naimplementovat algoritmus pro rozpoznávání a propojování pojmenovaných entit. Součástí tohoto cíle je také navrhnutí a vytvoření báze znalostí, která je v algoritmu použita. Vzhledem k omezenému množství dat pro jiné jazyky než pro angličtinu chceme, aby naši metodu bylo možné natrénovat na jednom jazyku a naučené parametry přenést na jiné jazyky (ve kterých není tolik trénovacích dat). Práce se skládá z popisu dostupnýchází znalostí, existujících metod, navrhnutí a implementace vlastní báze znalostí a metody pro propojování pojmenovaných entit.

Na několika variantách tradičního datasetu AIDA CoNLL-YAGO dosahuje implementovaný algoritmus nejlepších známých výsledků. Na vzorku českých anotovaných dat z datasetu PDT dosahuje algoritmus pomocí parametrů natrénovaných na anglickém CoNLL datasetu srovnatelných výsledků.